图书情报工作
第 63 卷 第 20 期 2019 年 10 月

学科热点概念的增长规律及属性分选研究

——以我国图书情报学领域为例

■ 高劲松 韩牧哲
华中师范大学信息管理学院 武汉 430079

摘要：［目的/意义］增长规律是信息和知识的动态特性，透过增长形态可以在一定程度上对知识元素的属性进行揭示，这对于知识组织和知识网络构建研究具有实际意义。［方法/过程］在对文献增长规律进行简要回顾后，以图书情报学近 25 年来热点概念的历年频次数据为基础构建数据集，进行曲线拟合；同时对拟合结果进行分解分析，进而根据逻辑函数的性质建立一种增长曲线形态测度指标，并进行实证分析。［结果/结论］通过曲线拟合和分解分析验证逻辑增长规律对信息或知识增长描述的有效性，实证分析表明提出的测度指标对学科热点概念的属性有良好的分选性，据此可将图书情报学的热点概念按象限划分为长效基础概念、短效老化概念和发展中热点概念三组。

关键词：热点概念 图书情报学 增长规律 逻辑曲线
分类号：G252
DOI：10.13266/j.issn.0252－3116.2019.20.006

1 引言

已发表文章的增长、老化和离散规律被视为标志科学文献发展的最根本规律，对信息情报增长、累积规律的探索与应用，一直以来都是图书情报学持续关注的重要课题之一。

在当前的情报学研究中，传统的以文献为计量单元的信息处理和信息组织工作已经很难满足新环境下的分析需求，人们倾向于突破文献单元的桎梏，对更细粒度的文献内容以及更深层次的知识单元进行分析与研究[1]。于是，传统的文献计量也逐渐向信息计量和知识计量转变[2]。知识计量的方法和理论体系建立在情报学的各种基本规律之上，而这些基本规律大多是以传统的文献计量方法推导出的，面向以内容和知识为单元的研究对象，以文献信息流为基础归纳总结出的规律是否适用，针对新的研究对象，这些规律是否有新的应用价值，也是值得关注的问题。对于科学文献而言，关键词、主题词、特征词等都是从文本内容中归纳、抽取出的，它们是对相关领域热点概念的词汇表示，对应文献的内容单元，这些词汇本身及其组合具有主题揭示作用，而对这些词汇的标识过程也可视为一种对构成知识元的知识要素的手工标记和提取工作。因此，从一个学科的宏观视角来看，由这些词汇组成的学科热点概念集合，也是构成该学科知识单元的基本元素集合，掌握学科领域发展历程中热点概念的更迭状况，可以把握学科主体的动态演化过程和趋势：肖明、李国俊等通过对 1998－2007 年国内情报学期刊论文的关键词词频进行统计，对目标时段国内情报学的研究热点及学科发展趋势进行了分析[3]；王连喜、曹树金从微观方面利用词频和 TF-IDF 方法提取重要关键词进行共词聚类分析，从全局方面运用 LDA 模型进行主题建模，二者结合对国内图书情报学和新闻传播学有关网络舆情研究的主题进行比较分析[4]；周鑫、蒋勋等综合运用词频变化率模型以及共现网络分析、多维尺度分析方法从研究热点、研究主题、研究范式 3 个方面对美国情报学研究的发展动向进行分析[5]；魏伟、郭崇慧等运用文本挖掘方法，对国务院政府工作报告（1954－2017 年）进行特征词筛选和时序聚类分析，归纳出历年来国务院工作报告中的 9 种特征词时间序列模式，揭示了建国以来的社会变迁状况[6]。上述研究充分证明了热点概念分析对于从知识层面把握领域动态状况的重要性，不过相关研究多是基于概念词频和

作者简介：高劲松（ORCID：0000－0003－0022－5923），教授，博士生导师，E-mail：jsgao@mail.ccnu.edu.cn；韩牧哲（ORCID：0000－0002－8474－4507），博士研究生。
收稿日期：2019－03－03 修回日期：2019－06－17 本文起止页码：51－61 本文责任编辑：徐健

时序分布状况对学科发展趋势进行考察,鲜有从热点概念本身出发,对其时序统计形态规律的验证和分析,事实上,这对于传统文献计量规律在知识层面的应用具有重要的理论意义。时序增长累积规律代表着文献信息流的动态特性[7],而热点概念的增长累积规律同样也指向目标学科领域随着时间发展的知识集中与离散现象,笔者对图书情报学近25年的学科热点概念进行年代、频次数据采集,并对筛选出的学科热点概念的增长累积规律进行统计和回归分析,试图探究传统的科学文献增长规律在知识层面的适用性,基于增长规律的数学性质建立分选指标,对不同学科热点概念增长形态的差异性进行分析,并对其在知识组织方面的实际应用进行尝试。

2 有关文献增长规律的研究

信息的增长规律研究最早是根据科学文献的增长规律总结出来的,20世纪40年代以来,以D. 普莱斯(D. Price)为代表的文献计量学家提出了一系列经典规律,用于描述文献信息的发展变化过程和分布状况。文献增长规律是后续信息增长规律研究的基础,文献增长规律中最具代表性的分别为指数增长规律和逻辑增长规律。

2.1 指数增长规律

指数函数是一种单调增长的无界函数,其最原始的现实意义是,在不受限制的理想情况下,某个系统的个体数量在一段时间之后会翻倍。

D. Price通过对数十种期刊和论文数量的增长特点进行统计分析,发现了文献随时间按照指数增长的规律[8]。据测算,正常的科学领域内的文献是按指数增加的,大约每隔10-15年时间增加一倍,年均增长约5%-7%。以文献数量为纵轴,时间为横轴,将不同年代的文献量在坐标系上逐点描绘并以平滑的曲线进行回归拟合,即能得到一条可以表征科学文献随时间增长的指数曲线,即“普莱斯曲线”。文献指数增长函数的数学表达式如下所示:

$$f(t) = ae^{bt}, a>0, b>0 \qquad 公式(1)$$

上式中,t是自变量,代表时间;$f(t)$为因变量,代表t时刻时的文献量;a为条件常数,即统计初始时刻的文献量;b为时间常数,即持续增长率。

由公式(1)可以继续推导出一个衡量文献增长速度的定量指标,即某一领域科学文献量翻倍所需时间,该周期记为T,计算公式如下所示:

$$T = ln2/b, b>0 \qquad 公式(2)$$

由于科学文献的增长速度不均匀,不同学科文献量翻倍的时间周期T并不相同,相关统计表明,有些学科平均几年就会翻一番,而一些学科则可能需要十几年。

随着相关学者对指数增长率探究的进一步深化,人们发现从较长的周期来观察,科学文献并不能一直保持指数增长的理想状态,即在不考虑长期保存限制的条件下,科学文献的累积虽然是恒增长的,但其增长速度会因主客观条件的限制而发生变化,即指数增长规律的失效问题,D. Price也注意到了指数增长规律的这一问题,并在其著作《小科学,大科学》一书中,进一步提出了科学文献的逻辑增长规律,并指出“指数型规律终将成为逻辑型”[7]。

2.2 逻辑增长规律

逻辑(logistic)函数是比利时数学家韦吕勒(P. F. Verhulst)在研究人口增长与资源承载量关系的问题时提出的数学模型,其曲线形态呈S型,因其丰富的逻辑推理含义,被称之为逻辑函数[9]。

科学文献的逻辑增长模型是前苏联科学学家纳里莫夫(B. Налимов)和弗莱杜茨(Г. Владуч)经过实验研究后提出的。他们发现科学文献的增长是分阶段的,每一阶段的增长模式并不相同,在科学文献累积的初始阶段,其遵循的是指数增长规律,不过随着时间的变化,其增长速度就会逐渐放缓,总体累积量最终会趋近于一个极限值。于是,科学文献的逻辑增长规律的增长速度呈先慢、后快、再慢,其累积曲线呈S型。科学文献的逻辑增长曲线惯用的数学表达式为:

$$f(t) = \frac{K}{1+ae^{-bt}}, a>0, b>0 \qquad 公式(3)$$

上式中,t是自变量,代表时间;$f(t)$为因变量,代表t时刻时的文献量;$K=\lim_{t\to\infty} f(t)$,即目标领域文献总体累积量的预期值;a、b为正参数。

逻辑增长是有限制的增长理论,在现实情况下,之所以不能维持无限制的指数增长,主要是受到系统环境阻力的限制。对于限制科学文献增长的环境阻力,B. Налимов将其归纳为物质条件、经济来源及作者智力等方面的因素。

对于较长的统计周期而言,逻辑增长规律对于文献增长累积状态的描述比指数增长规律更为贴切,也能更好地与生命周期理论相结合阐释其现实意义。不过,邱均平指出,逻辑增长规律“虽然克服了指数曲线‘发散性’缺陷,但也存在着‘有界’的局限”。在现实情况下,科技文献数量增长速度的减缓,并不完全意味

高劲松，韩牧哲.学科热点概念的增长规律及属性分选研究——以我国图书情报学领域为例[J].图书情报工作,2019,63(20):51－61.

着科学发展速度将会下降[7]。关于科技文献增长的相关统计中,也存在文献累积量在平台期再次突破,再度加速累积的现象,据此也有学者提出分阶段拟合的文献逻辑增长模型[10]。

2.3　关于文献增长规律的其他研究

上述两种规律都是描述科学文献增长的基本模型,近半个世纪以来,国内外相关学者一直在不断试图对其进行验证和改进,或提出新的增长模型,以便对科学文献的增长规律进行更精确地阐释。

其中,国外学者提出的比较有代表性的模型有:分级滑动指数模型[11]、超越函数模型、舍布增长模型[12]等文献增长修正模型;国内学者提出的模型主要包括:可变指数模型[13]、韩冬梅模型[14]、知识增长的动力学模型[15]、修正指数增长模型[16]、系统论模型[17]和考虑环境限制的知识增长模型[18]等。上述模型各有其合理性和缺陷,杨红、赵本东曾对国内的几种文献增长模型的优缺点进行评述,进而根据乘法原理修正指数增长模型,并最终推演出逻辑增长模型,据此论证了指数增长规律和逻辑增长规律的统一性,文章认为指数增长规律是逻辑增长曲线的特殊形式,科学文献的逻辑增长函数中的 K 值的实际意义是"研究对象领域的大小",在发生革命性变化时,K 值(原文用 B 表示)是可以继续提高的规模参数[19]。这样就合理解释了逻辑增长规律的"有界性"问题,不过上述模型的推演都是以科学文献的数量为研究对象的,一旦打破文献单元的限制,将研究对象细化到内容单元或知识单元,基于文献的逻辑增长规律是否有效、基于文献计量对逻辑增长规律现实意义的理解能否继续适用,都是值得进一步探讨的问题。

3　图书情报学热点概念增长累积规律

笔者以图书情报学近 25 年的学科热点概念为研究对象,通过考察这些热点概念的时序增长累积状况,根据统计指标进行横向的对比分析,并尝试通过分析各热点概念增长累积状态的差异,结合学科热点概念本身的语义信息把握近年来图书情报学的发展规律。

3.1　数据采集与预处理

笔者以中国知网(CNKI)数据库的公开汇编资源为数据源进行数据采集,采集对象为中文学术期刊数据库中呈现的主题词,这些主题词是 CNKI 从符合检索条件的论文关键词中自动筛选并整理出的高频词,这些高频词对符合检索条件的论文主题有较好的代表性和指向性,因此被作为主题分组依据方便用户对文献进行主题检索,系统同时提供了主题词的出现频次。考虑到数据来源代表性及数据处理规模的问题,笔者将数据采集的范围限定为被中文核心期刊数据库和 CSSCI 数据库检索的期刊。中文期刊全文数据库目前仅支持对 1992 年以后的数据进行采集,而 2018 年的数据不完整,因此本次检索的时间范围为 1992－2017 年。

数据采集和数据预处理工作是交替进行的,具体步骤如下:

(1)分年主题词采集和整理。在中文期刊数据库中,设定检索范围:文献分类目录－信息科技－图书情报与数字图书馆;设定文献来源类别:核心期刊 & CSSCI;在分组浏览选项中选取"主题"类目,分别对 1992 年至 2017 年共 25 年的数据进行检索,分年采集主题词,共得历年主题词 806 个;用 Excel 整理采集到的历年主题词表,设置条件格式,筛除 703 个重复主题词,得到 103 个不同的代表性主题词。

(2)主题词初次清洗。依据词意对上述 103 个不同的代表性主题词进行清洗,主要包含模糊主题词筛除,如"建筑物""各科经济"等由大量同类的不同主题词汇总而得的模糊主题词,其本身是没有专指性的词汇集合,不适合进行分析,故筛除;同义筛除,如"阅读推广"和"阅读推广活动",二者含义相似且增长累积趋势大致相同,保留"阅读推广"一词,其同义主题词的频次数据不做合并,直接筛除;共处理主题词 18 个,清洗后剩余主题词 87 个。

(3)主题词－频次矩阵建立。由于系统筛选出的不同年份的主题词在数量和内容上均不相同,笔者以初次清洗后得到的 87 个主题词为关键词(模糊),在中文期刊数据库中对其历年频次数据进行检索,检索范围为信息科技大类的核心期刊与 CSSCI 期刊论文,其实际意义代表在目标年份发表的以相关主题词(模糊)作为关键词的论文篇数;构建行为年、列为主题词的频次矩阵。

(4)主题词二次清洗。对主题词－频次矩阵进行整理,筛除非连续主题词,如"发文量"等主题词的历时频次数据不连续,存在多处 0 值,无法进行回归分析,故筛除;筛除数据量不足的主题词,如"数字人文"等主题词出现连续频次数据的时间很短,属于很难通过回归分析进行趋势判断的新生概念,故筛除;共处理主题词 8 个。

经过以上数据采集和预处理步骤,共得到 1992－2017 年代表图书情报学热点的主题词及其频次信息

79 组。

3.2　增长累积曲线拟合

之所以对各个热点概念的增长累积曲线进行拟合,而不直接选用历时频次数据的主要原因在于,增长累积曲线是一种单调递增曲线,相比拥有单峰甚至多峰的历时频次曲线而言,其曲线形态更简单,数据分析模型也更加成熟。

预处理后的数据是 79 个图书情报学热点概念在 1992 - 2017 年间的历时频次矩阵,在进行曲线拟合之前,需要对其作进一步处理:设定各个热点概念出现的起始年份为 t_0,并令任意 $t_0 = 1$;对同一热点概念,其后任一年的时间数据 $t_n = t_0 + n$。

对各个热点概念的历时频次数据进行累加,将任意热点概念每一年的统计频次记为 y,累加的初值即为该热点概念在起始年份的频次值,记为 $f(t_0) = y_0$;于是,同一热点概念在第 k 年的频次累加值 $f(t_n) = \sum_{k=0}^{n} y_k, k \leqslant n$。

将整理后的时间 - 累加值数据逐条导入数据分析软件 GeoGebra,进行双变量回归分析,其中时间 t 为自变量,各年累加值 $f(t)$ 为因变量,该软件自带多种常用的回归模型,本文实验根据相关热点概念的 $(t, f(t))$ 散点图预判曲线形态选择模型,模型选择标准为判定系数 $R^2 \geqslant 0.95$,若同一概念的增长累积曲线有两种或两种以上可选模型的话,则取 R^2 最高者。

按上述方法对 79 组数据逐项拟合,可知其中 61 组的拟合结果为逻辑曲线,11 组指数曲线,7 组三次多项式曲线。

上文对于科学文献的指数增长规律和逻辑增长规律的统一性问题已有所叙及,在一个领域的初期研究成果较少,可以用逻辑函数近似推导出指数函数形态,换言之,指数增长规律可以视为逻辑增长规律的特殊形态。而拟合得到的三次多项式如何加以解释,能否说明在文献或知识的增长累积过程中存在新的规律,就需要结合具体的热点概念,针对逻辑增长模型作进一步分解和剖析。

3.3　逻辑增长模型的分解

逻辑函数是一种增长趋势为"慢 - 快 - 慢"三段式的单调增函数,对逻辑函数(公式 3)求导,可得逻辑函数的速度函数,其表达式为:

$$v(t) = f(t)' = \frac{Kabe^{-bt}}{(1 + ae^{-bt})^2} \qquad \text{公式(4)}$$

继续对速度函数(公式 4)进行二阶求导,并令二阶导数为 0,可以计算出逻辑函数增长过程中的 3 个关键点,分别为速度函数的两个拐点和一个极值点。

其极值点的横坐标为:

$$t_H = \frac{lna}{b}$$

其拐点的横坐标分别为:

$$t_1 = \frac{lna - 1.317}{b}, t_2 = \frac{lna + 1.317}{b}$$

这 3 个关键点对应逻辑函数所反映的社会经济增长的始盛期、高峰期、盛末期;其中两个拐点又可以将其增长过程划分为渐增期 $(0, t_1)$、快增期 (t_1, t_2) 和缓增期 $(t_2, +\infty)$ 3 个阶段[20]。同时,有 $f(t_H) \approx 0.5K$,即当增长累积值达到预期规模一半的时候,逻辑函数的增长速度最快,此后其增速开始逐渐放缓。

下面结合"图书馆管理"这一热点概念的实际情况对逻辑曲线进行分解和分析。统计数据显示,"图书馆管理"在 25 年间共呈现 5 053 次,在统计初期正处于萌芽状态,在第 10 至 15 年间达到其研究的鼎盛期。对这项数据进行一次整体拟合、两次分段拟合。所得结果如图 1 所示:

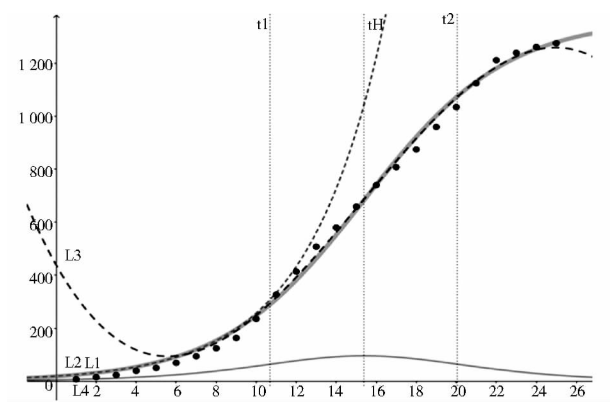


图 1　热点概念"图书馆管理"的拟合结果

图中共包含 1 条散点图和 4 条曲线,其中曲线 L_1 为热点概念"图书馆管理"的整体拟合曲线,其拟合结果为逻辑曲线,$R^2 = 0.996\ 9$,拟合效果优,$L_1: f(t) = \frac{1366.66}{1 + 74.32e^{-0.28t}}, (1 \leqslant t \leqslant 25)$。

曲线 L_2 为热点概念"图书馆管理"前半部分的分段拟合曲线,选取前 11 年的数据进行拟合,前 11 年是该热点概念的缓增期,其拟合结果为指数曲线,$R^2 = 0.979\ 5$,拟合效果好,$L_2: g(t) = 19.1e^{0.26t}, (1 \leqslant t \leqslant 11)$。

曲线 L_3 为热点概念"图书馆管理"后半部分的分段拟合曲线,选取第 11 至 25 年的数据进行拟合,可以与 L_2 分段拟合曲线的取值区间连续,其拟合结果为三

次多项式曲线，$R^2=0.9974$，拟合效果优，$L_3:h(t)=-0.32t^3+14.66t^2-134.84t+438.96$，$(11\leqslant t\leqslant 25)$。

曲线 L_4 为曲线 L_1 的速度曲线，$L_4:v(t)=f(t)'=\frac{28439.65e^{-0.28t}}{(1+74.32e^{-0.28t})^2}$，$(1\leqslant t\leqslant 25)$，图中 3 条垂直线为测算出的 3 个关键时间点的坐标，$t_1=10.6835$，$t_2=20.0906$，$t_H=15.3871$，分别对应 L_4 的两个拐点和一个极值点。

从图 1 可以看出，热点概念"图书馆管理"整体拟合所得的逻辑曲线，其渐增期曲线与指数曲线重合，快增期至缓增期前期的曲线则与三次多项式曲线两个极值点之间的 S 形部分走势趋近。

3.4 学科热点概念的逻辑增长机理分析

结合统计数据查证，上述 11 组拟合结果为指数曲线的热点概念，其逐年增长频次差值均为正值，至统计截止时尚未达峰值，这种特征对应逻辑增长的渐增期；而 7 组拟合结果为三次多项式曲线的热点概念，在统计初期已有较高的呈现频次，说明这些热点概念在统计初期既已步入快增期或缓增期。这 18 组数据在统计区间内都未能对其所对应热点概念的生命周期进行完整展示，不过从目前已展示的部分判断，其拟合结果均可视为逻辑曲线的分段形态。因此，上述 79 组数据的增长形态皆可用逻辑增长模型进行描述，我国图书情报学热点概念的增长累积亦服从逻辑增长规律，据此可进一步推断，逻辑增长规律不仅适用于描述有界的文献增长，也可用于描述无界的信息增长或知识增长。

进一步测算可知，在定义域$[t_1,t_2]$内，函数 $f(t)$ 所对应的值域约为$(0.21K,0.79K)$，即在整个逻辑增长的快增期，频次累积比例可占到预期规模的 58% 左右。而时间点 t_1 则大致可以视为指数增长形态与逻辑增长形态的分水岭，频次累积一旦达到预期规模的 21%，环境阻力就开始发挥显著作用，使得频次的增长累积无法继续保持指数增长势头，而逐渐转变为逻辑增长形态。

对本文所研究的热点概念而言，这种环境阻力并不能如科学文献增长一般归因于物质条件、经济来源及作者智力等方面的因素，这些热点概念的呈现是从群体性学术研究的成果中归纳而得，并不局限于某作者或作者群体的物质经济条件和意向选择。这些热点概念的增长累积趋势对应的是相关知识在相应时间段内的集中趋势，直接反映相关学科的发展状态及其需要解决的主要问题。不过，任何学科都存在其各自的发展轨迹和生命周期，一个学科在不同时代所面临的问题具有针对性和阶段性，在学科发展的历程中，很多阶段性问题会得到解决或者被更重要的问题所取代，在这些问题解决的过程中，以及科学技术和社会环境的发展变化的过程中，又会不断因新需求而催生新问题，整个学科的发展历程本质上就是不断提出问题、分析问题、解决问题的过程，这个使得科学研究不断规模化、精细化过程的直接表现就是知识不断累积的过程。因此，就本文而言，阻碍这些热点概念无限增长的直接原因就是其对应问题研究热度的减退或转移。导致这种热度减退或转移的因素较为复杂，主要是因科技环境和社会环境发展而导致的相关科研课题的解决、老化以及主要问题的转移。

4 基于逻辑增长规律的概念属性分选

对热点概念属性的辨识和分选是进行分类和组织的依据和前提，根据不同的需求，学科热点概念的属性可以从不同角度进行考察。国内外针对学科热点概念属性的专门研究都较少，但是在基于关键词或主题词对学科结构探测、发展趋势分析一类的研究中，依据一定属性对词汇进行分类是不可或缺的部分。词汇的属性[21]研究主要集中在自然语言处理领域，文献计量学中除了依据词汇在相应文献单元中的词频[22-23]、共现[24]、引用[25]情况对其属性进行分选的方法之外，近年来也有学者以词汇本身的语义属性作为分类依据，但相关的语义分类方法并不成熟[26]。基于上述属性分选方法进行概念分类，再通过静态或动态方法对概念本身所包含的信息进行分析，可以从一定程度上把握目标学科的结构及发展趋势。不过，动态发展的趋势本身也可以作为事物的一种属性，将趋势属性作为分类依据的研究主要集中在学科间或领域间的对比研究中。

对上文拟合所得的各热点概念的逻辑增长曲线在同一坐标平面上观察，不难发现，不同热点概念在增长累积形态上差异性与相似性并存，且遵循一定的规律。本节将上述热点概念的增长累积形态作为目标属性进行考察，结合逻辑函数的数学性质建立相应的形态测度指标，尝试对热点概念的增长累积形态属性进行分选，并结合概念本身阐述这种属性分选的现实意义。

4.1 逻辑增长形态测度指标的建立

增长累积规律符合逻辑函数的热点概念自出现以来的增长速度为先慢、后快、再慢的三段式，其速度函数的快增期对应相应概念在其整个生命周期之中的最

为盛行、相关研究最为集中的阶段，据上文测算，快增期的频次累积比例可占到预期规模的 58% 左右，所以，该阶段也可定义为相关热点概念的流行阶段。该流行阶段的时间跨度可记为 $\Delta t = t_2 - t_1 \approx 2.634/b$。

据统计，上述图书情报学热点概念的平均增长趋势也可用逻辑函数进行描述，其流行阶段并无实际意义，但可作为标尺衡量单个概念流行阶段的长短水平，其时间跨度记为$\overline{\Delta t}$，经过计算，图书情报学热点概念的平均流行阶段时长约为 14.633 年。该数值受到各热点概念预期规模的影响，不能直接代表图书情报学热点概念流行阶段时长的平均水平，对代表相关热点概念累积规模的 K 值进行分组加权，计算后所得的平衡时长约为 $0.8 \cdot \overline{\Delta t}$，数值约为 11.706 年，说明图书情报学热点概念流行阶段的平均时长约为 12 年。对于单个概念而言，若其流行阶段时长 Δt_n 大于 $0.8 \cdot \overline{\Delta t}$，则将其称为长效概念，反之，则称为短效概念。

除横向对比各热点概念流行阶段时间跨度的相对长度以外，对同一热点概念自身增长速度的衡量也可以从侧面反映相关热点概念的增长累积特性。逻辑曲线的速度曲线是以 0 为起点的单峰曲线，当 t 在(0，$2t_H$)的定义域内，速度函数的曲线为以 $t = t_H$ 为对称轴的正态曲线，当 $t = 2t_H$ 时，速度函数的值就无限趋近于 0，其对应逻辑函数的切线斜率也渐趋于 0。若不考虑多阶段增长的可能性，该阶段以后的逻辑曲线就成为无限接近 $f(t) = K$ 的直线，速度曲线呈现长尾效应。根据上文测算，在整个逻辑增长的快增期，频次累积比例可占到预期规模的 58% 左右，即某个热点概念在其流行阶段 Δt 期间的累积增长规模约为 $0.58K$。而对于任一符合逻辑增长规律的热点概念，从它产生到步入长尾期之前的整个增长时段的长度应为 $2t_H$，在这整个阶段中的累积增长规模则无限接近于预期规模 K。这样可以分别计算出相关热点概念在其流行阶段的平均增长速度 $v_h = \frac{0.58K}{\Delta t}$和整体平均增长速度 $\bar{v} = \frac{K}{2t_H}$，设该热点概念在流行阶段以外的平均增长速度为 v_s，则有 $\bar{v} = \frac{\Delta t \cdot v_h + (2t_H - \Delta t) \cdot v_s}{2t_H}$。理论上，在逻辑增长三阶段的累积规模相对于预期规模而言，其比例是固定的，因此，三阶段平均增长速度与整体平均增长速度的比例与三阶段的时间跨度在整体增长区间的占比直接相关。对这两种速度比值的测算可以反映增长曲线的形态，并有助于进一步探查不同增长形态背后的原因。当渐增期、快增期、缓增期三阶段时间跨度相等时，有 $2t_H = 3\Delta t$，此时的 $\bar{v}/v_h = \frac{1}{1.74} \approx 0.5747$，即当且仅当某一热点概念流行阶段的平均速度是整体平均速度的 1.74 倍时，其增长过程的渐增期、快增期、缓增期三阶段的时间跨度相同。以该值为标准，将所有的流行 - 平均速度比与该值做差，可以与流行阶段时间跨度的相对长度指标一起，判断相应热点概念的增长曲线形态。

笔者所分析的数据集中，有 18 组数据所得的拟合曲线为指数曲线和三次多项式曲线，上文结合实例对逻辑曲线进行分解，已经明确上述 18 组数据所对应的热点概念的增长累积规律符合逻辑增长规律，只是由于统计区间限制和发展阶段制约，使根据当前数据拟合得到的曲线不能展现逻辑增长的全貌。理论上可以结合灰色预测等方法先对相关数据样本进行补充，再推导其对应的逻辑曲线，但是，由于影响一个热点概念增长累积的现实因素较多且相对复杂，单纯根据现有数据进行预测补全，所得结果可能会有失客观，因此后续的试验和分析仅针对能直接拟合出逻辑曲线的 61 组数据进行。

根据上文分析，笔者拟以任一热点概念流行阶段的时间跨度与平衡计算后的平均流行时长的差值为横坐标、以任一热点概念的流行 - 平均速度比与三阶段时间跨度相同时的速度比的差值为纵坐标，构建直角坐标系，并在其中标记各个热点概念的位置。对任一热点概念而言，其对应点的坐标记为($\Delta t_n / 0.8\overline{\Delta t} - 1$，$\bar{v}/v_h - 0.5747$)，于是，该直角坐标系的 4 个象限便代表了 4 种不同的逻辑增长形态。见图 2。

第一象限，流行阶段的时间跨度大于平衡均值，且大于渐增期和缓增期的时间跨度，属于整体增长速度比较平稳的长效概念。

第二象限，流行阶段的时间跨度小于平衡均值，且大于渐增期和缓增期的时间跨度，属于整体增长速度比较平稳的短效概念。

第三象限，流行阶段的时间跨度小于平衡均值，且小于渐增期和缓增期的时间跨度，属于突增属性明显的短效概念。

第四象限，流行阶段的时间跨度大于平衡均值，且小于渐增期和缓增期的时间跨度，属于突增属性明显的长效概念。

4.2 图书情报学热点概念的分选与分析

为结合上述逻辑增长形态测度方法对各热点概念

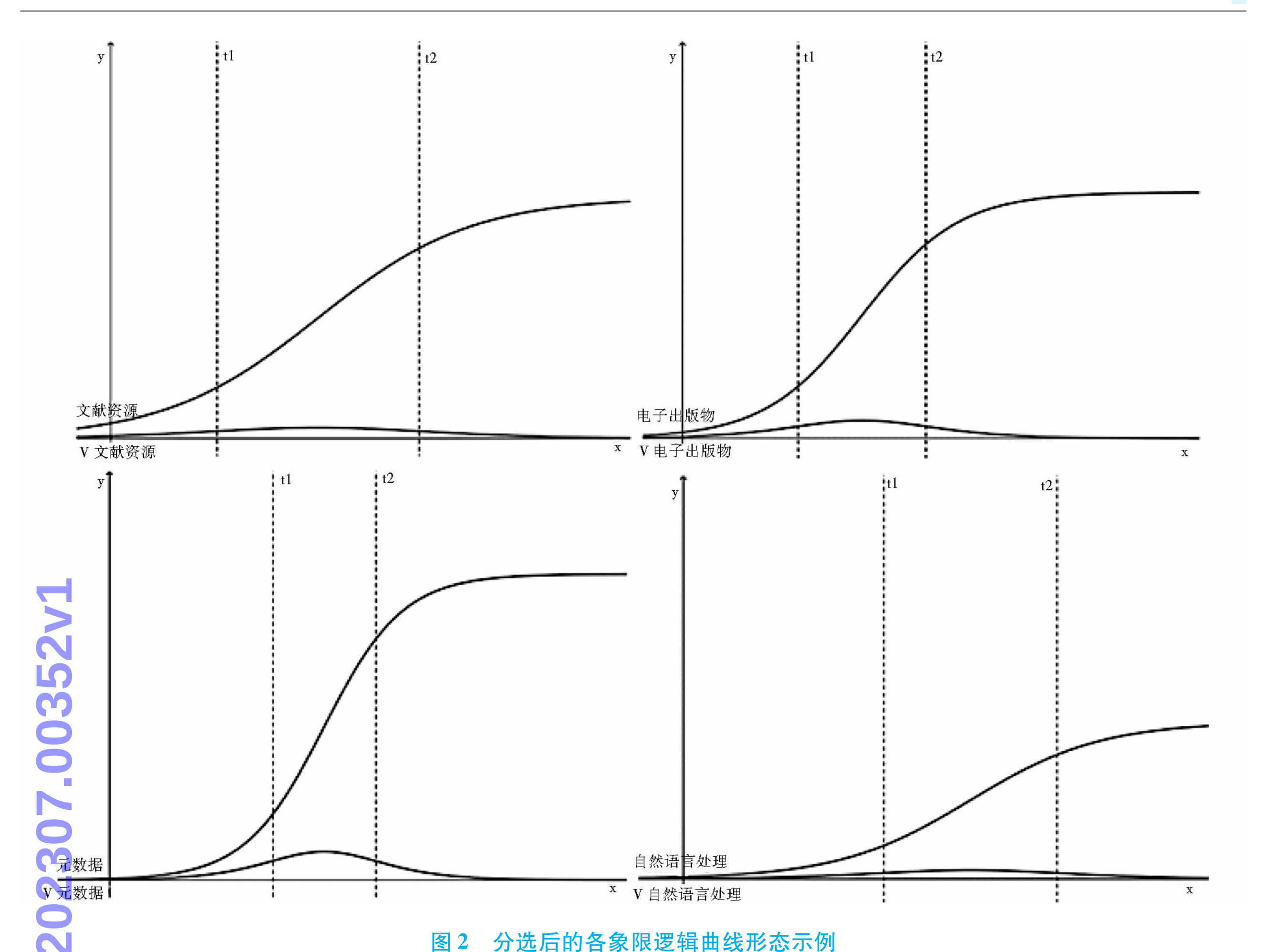


图 2 分选后的各象限逻辑曲线形态示例

的属性进行分选,在计算了上述 61 个热点概念所对应的坐标以后,笔者分别对其学科归属进行了追加统计,并将其初步划分为 3 种,分别为图书情报学内部概念以及外部的环境概念和交叉概念。这个类型划分是以相关热点概念的学科分布为基础,手工归类而得。若某热点概念在"图书情报与数字图书馆"这一类目的出现频次最高,则认为其属于图书情报学内部概念;反之,则将其归为图书情报学的外部概念,倘若其出现频次最高的学科同样归属于"信息科技"大类,则将其划为交叉概念,否则划为环境概念。统计可知,上述 61 个热点概念中,共有图书情报学内部概念 38 个,外部概念中的交叉概念 16 个,环境概念 7 个。

上述热点概念的学科归属及相关数据见表 1,将代表热点概念的各点在直角坐标系上展示,并根据其学科归属添加标识后的结果见图 3。

在一个学科领域中,这些热点概念及其组合代表着目标领域学者关注的集中点,其对应的是目标学科领域在一个时间段内需要解决的不同科学命题。一个学科领域中的科学命题也分为多个层次,从研究对象上来看,既存在基础性、全局性的命题,也有对方法论和学科分支命题的具体研究;此外,还有部分热点概念对应的并非研究对象本身,而是起到的是限定或修饰研究对象的作用,这些热点概念有些限定研究范围,有些则带有显著的时代特征。随着社会环境的发展变化,以及学科内部的理论创新和技术进步,同一学科领域在不同时间段所面临的主要问题和科研环境都会发生变化。因此,科学命题本身也有时代性及生命周期,这也是学科热点概念呈现逻辑增长的根本原因,由于不同的学科热点概念在学科领域中的属性和作用不同,它们的增长累积形态也存在差异。由图 3 可见,上文建立的逻辑增长形态测度指标对学科热点概念的属性具有较好的分选性。

从图 3 可以看出,上述热点概念的对应坐标主要分布在第一、三象限,第二象限的热点概念较少,第四象限仅有一个接近 y 轴的热点概念,不具代表性。由此可见,图书情报学热点概念的增长累积形态主要有 3 种,下面将结合相关热点概念的词义、统计特性及其学科归属对 3 个象限所包含热点概念的增长累积规律分别进行分析。

表 1 图书情报学热点概念的属性分选数据和结果

分布	$\bar{K}$	样本量均值	概念名称	学科归属	x	y	标识
第一象限	9437.635	24.95	公共图书馆	内部	0.875 107	0.005 324	●
			图书馆服务	内部	0.730 868	0.146 213	
			二次文献	内部	0.607 234	0.303 538	
			图书情报	内部	0.500 085	0.439 161	
			情报学	内部	0.406 33	0.168 354	
			图书馆学	内部	0.323 605	0.237 758	
			图书馆	内部	0.323 605	0.103 493	
			文献资源	内部	0.250 071	0.264 416	
			图书馆员	内部	0.250 071	0.296 502	
			情报工作	内部	0.184 278	0.237 403	
			竞争情报	内部	0.125 064	0.030 935	
			高校图书馆	内部	0.125 064	0.002 298	
			检索工具	内部	0.125 064	0.465 088	
			图书馆工作人员	内部	0.125 064	0.226 421	
			人工智能	交叉	0.125 064	0.050 073	○
			思维形式	环境	0.071 49	0.105 647	◆
			企业管理	环境	0.071 49	0.115 345	
			出版物	环境	0.406 33	0.270 312	
			美国	环境	0.323 605	0.214 57	
			中国	环境	0.250 071	0.111 348	
			财政管理	环境	0.184 278	0.059 786	
第二象限	2618.786	23.79	情报检索	内部	-0.021 68	0.467 783	●
			图书馆自动化	内部	-0.062 45	0.311 642	
			文献数据库	内部	-0.598 19	0.274 426	
			书目数据库	内部	-0.374 96	0.251 826	
			中图法	内部	-0.166 62	0.220 87	
			读者服务	内部	-0.196 38	0.213 953	
			数字图书馆建设	内部	-0.612 05	0.147 693	
			电子出版物	内部	-0.718 73	0.145 431	
			计算机	交叉	-0.718 73	0.137 468	○
			因特网	交叉	-0.249 96	0.110 744	
			信息高速公路	交叉	-0.134 57	0.076 474	
			高速信息传输网	交叉	-0.021 68	0.059 786	
			数据库	交叉	-0.437 47	0.038 661	
			知识经济	环境	-0.224 09	0.038 62	◆
第三象限	1798.105	21.24	社会网络分析	内部	-0.437 47	-0.014 02	●
			信息服务	内部	-0.134 57	-0.036 75	
			信息资源	内部	-0.224 09	-0.043 09	
			图书馆管理	内部	-0.196 38	-0.047 66	
			数字资源	内部	-0.391 86	-0.054 8	
			数字图书馆	内部	-0.338 2	-0.084 12	
			知识服务	内部	-0.391 86	-0.091	
			馆藏资源	内部	-0.021 68	-0.091 02	
			南大核心	内部	-0.499 97	-0.137 31	
			影响因子	内部	-0.196 38	-0.142 15	
			核心作者	内部	-0.062 45	-0.154 32	
			学科馆员	内部	-0.423 04	-0.136 79	
			网络环境	内部	-0.437 47	-0.162 48	
			图书馆联盟	内部	-0.488 61	-0.219 1	
			被引次数	内部	-0.357 11	-0.244 26	
			移动图书馆	内部	-0.691 76	-0.225 31	
			云服务	交叉	-0.673 89	-0.091 36	○
			信息检索	交叉	-0.249 96	-0.096 71	
			互联网络	交叉	-0.540 79	-0.107 8	
			云计算	交叉	-0.631 13	-0.108 09	
			虚拟现实	交叉	-0.318 14	-0.074 39	
			元数据	交叉	-0.391 86	-0.156 38	
			语义网	交叉	-0.451 19	-0.275 42	
			知识管理	交叉	-0.437 47	-0.179 43	
			用户行为	交叉	-0.391 86	-0.284 7	
第四象限	1088.780	25.00	自然语言处理	交叉	0.022 786	-0.056 18	○

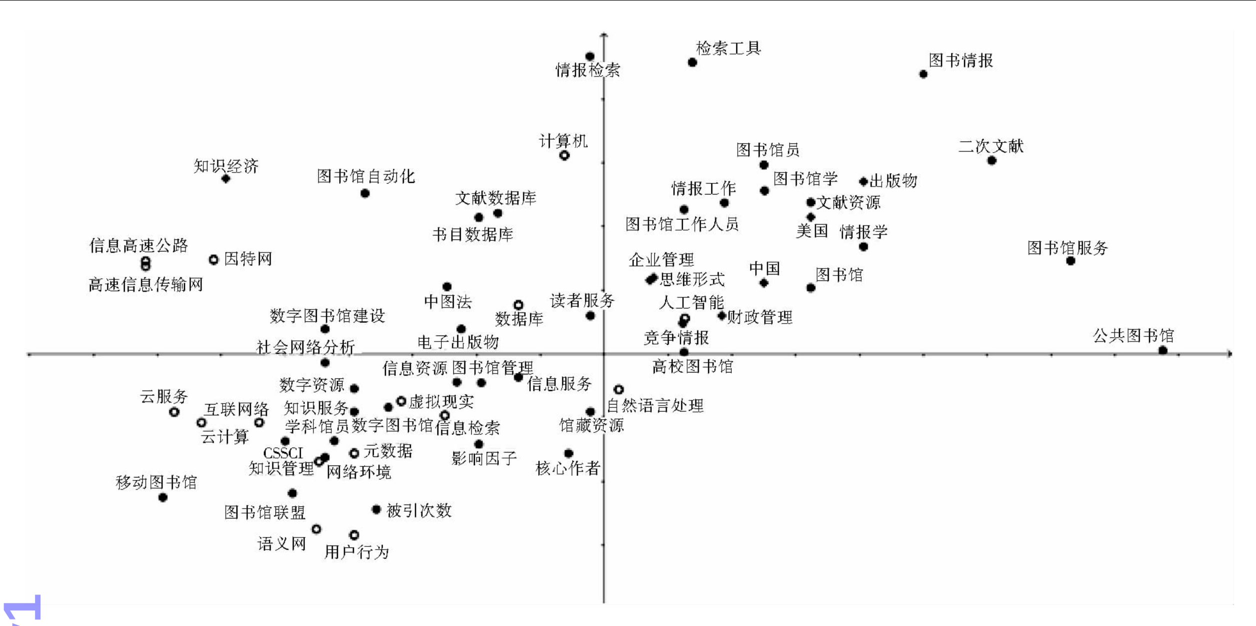


图3　图书情报学热点概念的属性分选结果分布

第一象限，对应增长速度平稳的长效概念。本文数据集中的长效概念在学科归属上主要是图书情报学内部概念和外部概念中的环境概念，该象限仅有一个交叉概念。分布于第一象限的各热点概念在该学科领域持续保持着较高的影响力，在较长的时间段内都保持着较平稳的研究热度。结合词义分析可知，分布于该象限的内部概念多具有基础性和全局性，图书情报学的学科主体、研究对象、情报信息的主要载体等基础性热点概念都在其中，涉及这些基础性概念的问题是伴随着这个学科领域一直存在的，而且随着学科的发展而不断更新。分布于第一象限的外部概念主要是环境概念，这些概念本身并不属于图书情报学，其对应的是近年来的社会经济环境热点，涉及到社会科学的各个方面，这些热点本身具有时代特征，且影响范围广、持续时间长，相对于图书情报学这个单一学科领域而言，这些随着社会经济环境的发展而变化的环境概念的存在具有普遍性和持续性。由于第四象限仅有的热点概念不具有代表性，因此可以将第一象限的热点概念统称为长效概念，这些长效概念在短时间内的增长速度比较平缓，K 值普遍较大，大部分长效概念至本文统计时仍呈现平稳的增长态势，由于拟合方法得到的曲线方程会受到数据样本量和统计区间的限制，有理由怀疑其中部分长效概念的实际增长规模更大且流行阶段的时间更长。

第二象限，对应增长速度平稳的短效概念。分布于第二象限的热点概念增长速度在本文统计区间内已普遍出现颓势，其中部分热点概念已经步入逻辑增长的长尾期。该象限的热点概念主要是内部概念和外部概念中的交叉概念，另有一个环境概念是“知识经济”。结合这些热点概念的统计数据可知，第二象限的概念已经出现了不同程度的老化，主要可分为以下几种情况：①概念指向信息本身的老化，例如，传统依托书报期刊等纸媒的文献服务在信息时代已逐渐转变为依托数字信息和线上数据平台的信息服务，此前通常用来指代图书馆服务对象的读者，也泛化为信息用户，由此一来，“读者服务”这一概念被提及的频次就越来越少；②概念称谓形式的老化，例如早期多以 Internet 的音译名词“因特网”指代互联网，随着称谓逐渐规范化，“因特网”这一概念在形式上也统一到互联网中，从而鲜被提及；③由外部概念基础化所致的老化，例如，在世纪之交，“计算机”和“数据库”在多个学科领域开始崭露头角，与此密切相关的数字图书馆建设工作成为图书情报领域当时研究的重点，经过十几年的发展，计算机科学已经成为图书情报学的主要交叉学科之一，对数据库及数字图书馆建设的研究与实践也逐步深入、细化，之前较为宽泛的理论研究已经转变为如今的基础性成果，这些热点概念也已成为一个个学科分支，有着更精细化的学科结构，对其本身的研究也就逐渐内化其中了。综上，第二象限的热点概念主要是一些趋于老化的短效概念，在统计期内，用于拟合的数据基本能够覆盖这些热点概念的整个生命周期，在统计期末，分布于第二象限的热点概念的预期规模 K 值已普遍趋于稳定。

第三象限，对应突增属性明显的短效概念。结合统计数据观察，发现大部分热点概念虽然在统计期内，其增长速度已经出现峰值，但在统计截止时仍保持着

增长态势,且第三象限的新增热点概念较多,部分概念出现年份较晚,用于拟合的数据平均仅有 21.24 年。分布于第三象限的热点概念中主要包含涉及图书情报学具体研究领域和学科分支的对象、方法、理念和技术,以及与计算机科学、管理学等学科领域的交叉概念。这些概念含义具体,对应的都是近年来讨论较多的研究热点,有些概念自学科自身的理论及技术创新中产生,有些自相关学科或交叉学科中借鉴,这些概念自出现起就引起图书情报学者的关注,故而,在这些概念的流行阶段,其表现出的增长速度非常快,突增属性明显。这些热点概念有些至今仍保持着较高的研究热度,与第三象限概念相关的"数字图书馆发展""图书情报学期刊发展""五计学""图书馆阅读推广""图书馆知识服务"等学科主题频繁出现在近年来国家社科基金项目的历年选题指南中[27]。不过一些更前沿的研究热点的出现也使得研究热点不断更新,传统概念的研究热度不断被新兴概念分散,其历时增长速度会逐渐放缓,有些概念会逐渐基础化转变为长效概念,有些则可能被新概念取代而趋于老化。从统计数据上看,分布于第三象限的热点概念普遍存在拟合样本量不足的现象,且 GeoGebra、SPSS 等主流的数据统计分析软件对逻辑函数的拟合都是根据终值推断 K 值,进而通过线性化运算得出增长函数,这种拟合方法适用于对数据样本量能够覆盖整个增长周期的函数拟合,但对于这些尚处于增长阶段的函数拟合可能会出现偏差。这意味着目前针对第三象限热点概念拟合出的增长函数并非稳态,随着时间的推移,数据样本量的增加,其增长函数会发生变化,对应的统计特征也会改变。第三象限只是当前时间节点上表现出的过渡状态,随着相关热点概念的发展成熟,其最终会稳定于第一象限或第二象限。这些过渡状态的增长函数及相应的测度指标并非无意义,其表现出的良好的分选性对确定热点概念的属性和状态具有重要的现实意义和实际的应用价值。

5 结语

笔者在回顾此前学者关于文献的逻辑增长规律和指数增长规律的基础上,以图书情报学近 25 年来的热点概念的历年频次数据为基础构建数据集,通过曲线拟合,并对逻辑曲线的图形进行分解分析,验证了逻辑增长规律对信息增长或知识增长描述的有效性。进而,根据逻辑函数的性质建立了一种增长曲线形态测度指标,实证分析表明其对学科热点概念的属性和发展状态有良好的分选性,并据此对各热点概念分组,分别为:第一象限的长效基础概念、第二象限的短效老化概念和第三象限的发展中热点概念。

笔者所分析的学科热点概念是构成该学科相关知识的代表性元素,这些热点概念与其他构成知识的元素一起,通过一些关联关系加以组合,构成了目标学科的诸多科学命题及其对应的知识。这些知识通过关联组合,构成了目标学科的知识网络,这个知识网络就是该学科的客观知识结构在现实中的映射。学科发展的不同阶段所面临的主要问题不同,而且不同时段内的学科热点也在不断变化更新,学科热点概念表现出的这种历时性新陈代谢事实上正是促进学科不断发展的推动力。一个学科的热点概念可能有着不同的属性,其自身发展也不均衡,在学科知识的构成上,不同概念的组合可能也会遵循一定的属性搭配,这种属性搭配可能进一步揭示构成知识的关联关系,并有望以此为突破口继续研究知识元的分组与组织以及知识网络构建的问题;而处于不同发展阶段的概念可能按照固定的搭配形式相互组合,从时序上加以梳理,可以进一步掌握一个学科领域知识的演化过程和动力机制。笔者针对热点概念的增长累积规律进行验证,并据此提出学科热点的属性分选指标,实是为进一步更深层次的知识组织研究做铺垫。

本文还存在一些值得继续研究和改进之处,文中对逻辑增长函数进行了分解,从图形上直观地解释了拟合结果中存在指数增长和三次多项式增长的原因,并据此论证了指数增长和逻辑增长的统一性,不过并未尝试根据 3 种增长函数的关联写出所有热点概念的逻辑增长函数,以致于在后续的试验和分析中舍弃了 18 组数据,这一方面是出于对数据精确程度的考虑,另一方面也实际受限于单一的函数拟合方法。此外,通过分析逻辑拟合结果对热点概念的增长和老化状态进行区分并不合理,笔者提出的测度指标之所以能对概念增长状态进行区分,原因在于第三象限的曲线拟合结果受限于样本量普遍未达到稳定状态,拟合结果本身带有一定的过渡性和阶段性,换言之,从整体上看,流行阶段的时间跨度大于渐增期或缓增期的时间跨度,可能是一个成熟的可以拟合出固定逻辑曲线的热点概念的固定属性,不过这一猜想可能需要更多样本进行验证。

参考文献:

[1] 宋艳辉, 王小平. 从文献单元、信息单元向知识单元的嬗变[J]. 科研管理, 2015(1):459-464.
[2] 赵蓉英, 魏明坤. "五计学"在我国的发展演进分析[J]. 现代情报, 2017(6):157-161,169.

[3] 肖明，李国俊，杨楠. 基于词频分析的国内情报学研究热点(1998－2007)[J]. 情报杂志，2009，28(8):21－25.
[4] 王连喜，曹树金. 学科交叉视角下的网络舆情研究主题比较分析——以国内图书情报学和新闻传播学为例[J]. 情报学报，2017(2):53－63.
[5] 周鑫，蒋勋，陈媛媛. 词频变化率模型视域下美国情报学研究发展动向分析[J]. 情报科学，2017(4):169－175.
[6] 魏伟，郭崇慧，陈静锋. 国务院政府工作报告(1954－2017)文本挖掘及社会变迁研究[J]. 情报学报，2018，37(4):70－85.
[7] 邱均平. 信息计量学(二)第二讲文献信息增长规律与应用[J]. 情报理论与实践，2000，2(23)：153－157.
[8] 普赖斯. 巴比伦以来的科学[M]. 任元彪,译. 石家庄：河北科学技术出版社，2010.
[9] VERHULST P F. Mathematical investigations on the law of population growth[M]// Annotated readings in the history of statistics. Berlin：Springer，2001:69－75.
[10] 王长安. 我国科技期刊逻辑增长模型的研究[J]. 情报科学，2004，22(5):554－557.
[11] RESCHER N. The limits of science[M]. Pittsburgh：University of Pittsburgh Press，1999.
[12] 邱均平，王宏鑫. 20世纪文献计量学发展的层次分析[J]. 高校图书馆工作，2000，20(4):1－8.
[13] 苏宜. 可变指数的文献指数增长规律[J]. 情报杂志，1996(5):11－13.
[14] 韩冬梅. 一个描述科技文献增长规律的新模型[J]. 情报理论与实践，1997(5):280－282.
[15] 卢贵武，韩冬梅. 科技文献增长的动力学模型[J]. 情报学报，1997(S1):353－356.
[16] 靖培栋，康仲远. 关于科技文献增长的数学模型[J]. 情报学报，2000，19(1):90－96.
[17] 王宏鑫. 文献增长经典模型的系统论统一[J]. 情报杂志，2003，22(5):2－3.
[18] 林辉，林伟. 科学文献的增长规律和老化规律及其新的一般模型[J]. 情报杂志，2010，29(4):22－25.
[19] 杨红，赵本东. 科技文献之逻辑增长模型的推演[J]. 首都师范大学学报(自然科学版)，2011，32(4):18－23.
[20] 崔党群. Logistic曲线方程的解析与拟合优度测验[J]. 数理统计与管理，2005，24(1):112－115.
[21] 杨宏进. 基于关键词检查属性指标的方法研究[J]. 统计研究，2013，30(8):10－16.
[22] 周鑫，陈媛媛. 关键词词频变化视角下学科研究发展趋势分析——以国内情报学研究为例[J]. 情报杂志，2016，35(5)：133－140.
[23] 奉国和，孔泳欣，肖洁琼. 基于加权关键词的领域热点与趋势分析新方法[J]. 图书情报工作，2018，62(18):102－109.
[24] 刘晓波.我国图书馆学研究热点及趋势:基于关键词共现和词频统计的可视化研究[J].图书情报工作,2012,56(7):62－67.
[25] 韩牧哲. 基于引文网络的知识扩散研究[D]. 曲阜：曲阜师范大学，2016.
[26] 刘自强，王效岳，白如江. 语义分类的学科主题演化分析方法研究——以我国图书情报领域大数据研究为例[J]. 图书情报工作，2016，50(15):76－85.
[27] 全国哲学社会科学工作办公室. 国家社科基金项目课题指南[EB/OL]. [2018－10－16]. www.npopss-cn.gov.cn.

作者贡献说明：
高劲松:论文指导、撰写与修改；
韩牧哲:数据采集、实验和论文撰写。

Validation on the Growth Law and Attribute Sorting of Field Hot-concepts: Taking Domestic Library and Information Science as an Example

Gao Jinsong　Han Muzhe
School of Information Management, Central China Normal University, Wuhan 430079

Abstract: [**Purpose/significance**] As a dynamic characteristic of information and knowledge, the growth law can reveal the attributes of knowledge elements to a certain extent, which has practical significance for the study of knowledge organization and knowledge network construction. [**Method/process**] After a brief review of the law of literature growth, this paper constructs a data set based on the frequency data of the hot-concepts in Library and Information Science in the past 25 years, carries out curve fitting, and decomposes and analyses the fitting results, and then establishes a growth curve shape measurement Indicator according to the nature of the logical function, then an empirical analysis has been carried out. [**Result/conclusion**] The validity of the description of the growth of information or knowledge by the law of logical growth is verified by curve fitting and decomposition analysis. The empirical analysis shows that the measurement index proposed in this paper has a good sorting ability for the attributes of the hot-concepts of disciplines. Based on this, the hot-concepts of Library and Information Science can be divided into three groups according to the quadrant: long-term basic concepts, short-term aging concepts and developing hot-concepts.

Keywords: hot-concepts　library and information science　growth law　logistic curve